Technical Report No. 32-705

The Application of Extreme-Value Theory to Error Probability Estimation in the Ranger Block III Command Detector

Edward C. Posner

M. Easterling

M. Easterling, *Chief*
Communications Systems Research

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

January 15, 1965

Prepared Under Contract No. NAS 7-100
National Aeronautics & Space Administration

CONTENTS

TABLE

FIGURES

ABSTRACT

This Report presents an application of Gumbel's extreme-value theory to the estimation of low error probabilities in certain types of communications systems such as the *Ranger* Block III Command Detector. The systems considered all involve threshold detectors in a binary system. By taking the maxima of a large number of successive independent samples, and then by taking a large number of these maxima themselves, the two parameters of the Gumbel distribution are estimated using large-sample theory. The variance of the error-probability estimate is found, as well as a confidence interval. The savings obtained by using the Gumbel method, as opposed to the simple monitoring of errors, is shown to be large. Finally, by applying a goodness-of-fit test to the voltage data of an actual *Ranger* Block III Command Detector, it is shown that the Gumbel distribution does indeed hold.

I. INTRODUCTION

The statistical theory of extreme values (Ref. 1 and 2) applies to the problem of estimating low error probability whenever errors are caused by large fluctuations of a random quantity such as a voltage. The basic idea is to use a short stretch of fluctuation data to estimate how long it would take for a random fluctuation to be so large as to cause an error. The technique may be used whenever one must identify very low error probabilities, or whenever the time or expense of estimating these probabilities by counting the occurrences of errors is prohibitive. To date, the applications of this technique have been in fields such as civil engineering and actuarial science, rather than in communication theory. However, a system is being designed for using this method in testing *Ranger* Block III Command Detectors, so that faulty detectors will not be put aboard the spacecraft; those detectors that give too high an error probability will be rejected.

Gumbel's extreme-value distribution functions are of the form $\exp-(\exp-(\alpha(x-u)))$, where α and u are positive parameters. This distribution is the asymptotic distribution, as $n\to\infty$, of the extreme positive value among n independent random variables x_i which are chosen from a distribution of exponential type on the right, if α and u are correctly chosen. "Exponential type

on the right" means that the distribution $F(x)$ has, for x greater than some x_0, a density function $f(x)$ which is differentiable and not zero, such that

$$\lim_{x\to\infty} \frac{f(x)}{1-F(x)} = -\lim_{x\to\infty} \frac{f'(x)}{f(x)}.$$

"Asymptotic distribution as $n\to\infty$" means the following: a sequence of distribution functions $G_n(w)$ is said to converge to the asymptotic distribution $G(w)$ if $\lim_{n\to\infty} G_n(w) = G(w)$ for all w.

The definition of asymptotic distribution applies to the extreme-value situation as follows. For each n, the distribution of the maximum of n independent samples from $F(x)$ has associated with it positive parameters α_n and u_n, whose definition need not concern us here. Let $z = \max x_i$ $(1 \leq i \leq n)$ and consider the random variable $w = \alpha_n (x-u_n)$. Then the asymptotic distribution holds in the following sense. Let G_n denote the distribution of w. The statement is that the distribution $G_n(w)$ converges to $\exp-(\exp-(w))$.

The parameters α_n and u_n do not converge to α and u. However, for given fixed n, the parameters α_n and u_n (written α and u from now on) are unknown in general, and so must be estimated. Of course, α and u are determined from F and n, but the point is that F is not known; in fact, if F were known, the extreme-value theory would not be needed. One is willing to assume, however, that the unknown F is of exponential type, since this assumption is satisfied for distributions having right-hand tails qualitatively like a normal or negative exponential distribution.

Now consider a "threshold detection" scheme in which a bit is called a "1" if the output of the detector rises above a certain level, and a "0" if it is below that level. We assume that the bit rate is slow enough so that deviations in different bits are independent random variables, as is often true in a well-designed system, and is true in the *Ranger* Block III Command Detector. We examine a run of $M = Mn$ successive bits with, for example, a "1" as the transmitted symbol. We look at the maximum of the (negative of) this deviation in each block of n bits. If n is large enough, the deviations would have the extreme-value distribution with unknown parameters α and u. Then we estimate α and u from the N samples of the extreme-value distribution.

Using α and u, we can then readily estimate the error probability. First, we estimate the probability that the maximum out of n exceeds the threshold. Simple algebra transforms this probability into the probability that a given single observation exceeds the threshold; for small error probabilities, we just divide by n. We arrive at an estimate of the probability of the error "1"→"0".

A similar estimate is used to find the error probability of the "0"→"1" transition. However, in the *Ranger* Block III Detector, the threshold is set asymmetrically, so that the "0"→"1" transition can be assumed not to occur. Thus, in such cases, assuming that a "0" and a "1" are equally likely to be transmitted, the error probability is ultimately estimated as one-half the estimate of the probability of the "1"→"0" transition.

In a typical case, we used $M = 3000$, $n = 100$, $N = 30$. That is, extremes were taken from successive 100 samples, giving 30 independent samples from an extreme-value distribution. As we shall see, to obtain with equal confidence an error-probability estimate using the method of counting error occurrences, the number of samples required would be ten or more times as large as the number we have used. It is this difficulty of obtaining sufficient data by classical error-estimation procedures, coupled with the failure of the normal distribution to adequately explain the distribution of voltage at an individual sample, that accounts for the usefulness of extreme-value theory in its application to the *Ranger* Block III Command Detector.

Previous uses of extreme-value theory have not been to estimate error probability. Rather, the method has been used to determine, for example, how large a dam to build to cope with maximum floods. In previous applications, the "threshold" was to be set after, rather than before, determining the maximum "voltage." In such previous applications, "threshold" would correspond to the size of dam to be built, and "voltage" would correspond to the height of water in the flood. In the present application, we are given a "dam" already built, and we must examine a short record of "floods" before paying for the dam. Thus, the model is somewhat different.

II. MAXIMUM-LIKELIHOOD ESTIMATORS

The best kind of estimator to use for general purposes is a maximum-likelihood estimator. These estimators are asymptotically unbiased and yield asymptotically minimum mean-square estimation error. Furthermore, maximum-likelihood estimators have an asymptotic joint normal distribution, whose means are the true means and whose covariance matrix is readily calculable. Thus, confidence intervals can be constructed for a single parameter without knowledge of the other parameters. This ability to obtain confidence intervals for a single parameter is a property of jointly normal estimators not shared by other kinds of estimators in general (Ref. 3, Chapter 10). The sample size N that one has to deal with in some of the applications can be as small as 20 or 30, so that the applicability of the asymptotic theory of maximum-likelihood estimators can be questioned. But the decided advantages of using the asymptotic theory are so great that we adopt the principle of maximum-likelihood and assume the asymptotic joint normal distribution.

To obtain the maximum-likelihood estimators, we first change parameters in the extreme-value distribution $F(x) = \exp -(\exp -(\alpha\,(x-u)))$ to a set of parameters more suited to our purposes. Namely, we are interested not so much in α or u as in β, the probability that a random variable distributed according to $F(x)$ exceeds the known threshold x_0. This β is thus defined as $F(x_0) = \exp -(\exp -(\alpha\,(x_0-u))) = \beta$. We now write F in terms of α and β, rather than in terms of α and u.

Since $\beta = \exp - (\exp - (\alpha(x - u)))$, $\log \beta = - \exp -(\alpha(x_0 - u))$. Let us define $\nu = \alpha(x_0 - u)$ and estimate ν; $\beta = \exp -(\exp -(\nu))$. Then

$$F(x) = \exp -(\exp-(\alpha(x - x_0) + \nu)) \tag{1}$$

is a Gumbel distribution, with unknown parameters α and ν. We now obtain the maximum-likelihood estimators of α and ν.

If the N samples from F are $x_1, x_2, \cdots, x_N$, then the likelihood function for the given sample of N is

$$L(x_1, \cdots, x_N; \nu, \alpha) = \alpha^N \exp -(\alpha \sum (x_i - x_0) + N\nu) \times \exp -(\sum \exp -(\alpha(x_i - x_0) + \nu)), \tag{2}$$

since the density function $f(x) = dF(x)/dx$ is given by

$$f(x) = \alpha \exp -(\alpha(x-x_0)+\nu) \exp-(\exp-(\alpha(x-x_0)+\nu)). \tag{3}$$

To find the α and ν maximizing equation (Eq. 2), we differentiate the logarithm of Eq. 2:

$$\log L = N \log \alpha - N\,\alpha\,(\bar{x} - x_0) - N\nu - \sum \exp -(\alpha(x_i - x_0) + \nu), \tag{4}$$

where $$\bar{x} = \frac{1}{n} \sum x_i .$$

Then

$$\frac{\partial \log L}{\partial \alpha} = \frac{N}{\alpha} - N\,(\bar{x} - x_0) + \sum (x_i - x_0) \exp -(\alpha(x_i - x_0) + \nu), \tag{5}$$

$$\frac{\partial \log L}{\partial \nu} = - N + \sum \exp -(\alpha(x_i - x_0) + \nu).$$

The likelihood equations to be solved for $\hat{\alpha}$ and $\hat{\nu}$, the required maximum likelihood estimators, are obtained by setting $(\partial \log L)/(\partial \alpha) = (\partial \log L)/(\partial \nu) = 0$ in Eq. 5.

The likelihood equations (Eq. 5) do not have a closed-form solution, so a numerical technique must be used to find $\hat{\alpha}$ and $\hat{\nu}$. Newton's method is useful, but a good 0^{th} approximation must be obtained before the method will work.

One way of obtaining first estimates graphically will be briefly described here; a numerical approach is expounded in Ref. 2, p. 226. We use extreme-value probability paper, which is probability paper so constructed that when an extreme-value distribution function is plotted it becomes a straight line (Ref. 2, p. 34). Such paper is very useful in this work, and is available from a commercial supplier.[1] The vertical scale is linear in x, but the horizontal scale is proportional to $-\log(-\log F(x))$. We order the x_i to obtain the order statistics $x_{(m)}$, where $x_{(m)} \geq x_{(m+1)}$, $1 \leq m \leq N$. The value of $F(x)$ to be associated with $x_{(m)}$ is $(N + 1 - m)/(N + 1)$.

[1] Technical and Engineering Aids to Management, Lowell, Mass.

We then fit a straight line to the data by any means, visual or otherwise, to obtain the first approximation to $\hat{\alpha}$ and $\hat{u}$. Thus, the slope of the line is the first approximation $\hat{\alpha}_0$ to $\hat{\alpha}$ and the vertical intercept is the first approximation $\hat{u}_0$ to $\hat{u}$. Then the first approximation $\hat{v}_0$ to $\hat{v}$ is $-\exp-(\hat{\alpha}_0(x_0 - \hat{u}_0))$. Newton's method for solving simultaneous transcendental equations is then applied until sufficiently good approximations to $\hat{\alpha}$ and $\hat{v}$ are obtained.

Thus, consider the data of Table 1, obtained from a 5-min run of a particular *Ranger* Block III Command Detector. Since ten samples that are still independent can be taken per second, $M = 3000$. For n, 100 samples were used. Thus $N = 30$. The data is in arbitrary units in which the threshold x_0 was 955, just beyond the highest x_i observed. The graph of Fig. 1 was obtained, with the line corresponding to the Gumbel distribution fitted by eye. Notice how good the fit is; (we shall say more about this in Part V). The values of $\hat{\alpha}_0$ and $\hat{v}_0$ obtained from the straight line were $\hat{\alpha}_0 = 8.77 \times 10^{-3}$, $\hat{v}_0 = 3.46$. The likelihood equations appear formidable, but the visual fit was sufficiently good that the likelihood equations were solved with a desk calculator, using only two iterations. The final values obtained were $\alpha_0 = 8.76 \times 10^{-3}$, $\hat{v} = 3.52$. Then $\hat{\beta} = \exp-(\exp-(v)) = .971$, and $1 - \hat{\beta} = 2.9 \times 10^{-2}$ is the maximum-likelihood estimate for the probability that the maximum in 30 of the x_i exceeds x_0.

Table 1. The thirty ordered extreme deviations $x_{(m)}$

m	$x_{(m)}$
1	948
2	881
3	837
4	820
5	793
6	758
7	750
8	732
9	724
10	720
11	697
12	650
13	645
14	642
15	623
16	604
17	591
18	582
19	570
20	561
21	540
22	520
23	503
24	489
25	477
26	460
27	450
28	442
29	423
30	400

Now let $\hat{e}$ be the maximum-likelihood estimate of the probability e of making the "1"→"0" error when a "1" is sent. We have $(1 - e)^n = \hat{\beta}$, so $\hat{e} \approx (1 - \hat{\beta})/n = (2.9 \times 10^{-2})/100 = 2.9 \times 10^{-4}$. Letting p be the error probability for the particular asymmetric detector being studied, we observe that $p = e/2$; so if $\hat{p}$ is the maximum-likelihood estimator of p, we have $\hat{p} = \hat{e}/2$. Thus $\hat{p} = 1.5 \times 10^{-4}$ is finally the maximum-likelihood estimate of the error probability using that particular detector under those particular environmental conditions.

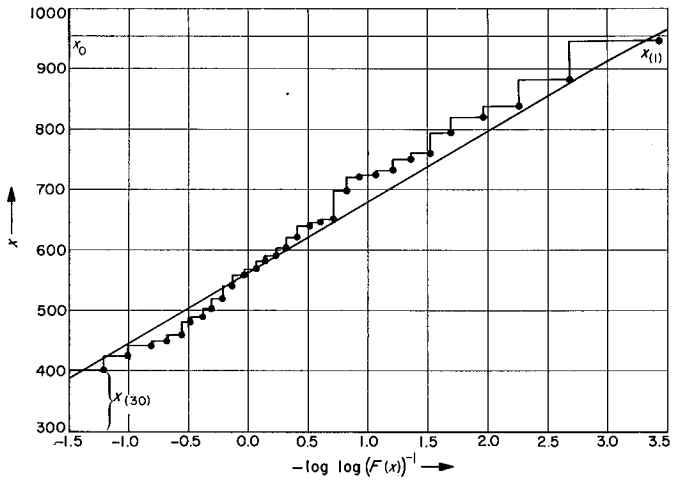

Fig. 1. Table 1 plotted on extreme-value paper

III. CONFIDENCE INTERVALS

Confidence intervals for $\hat{\nu}$ are obtained by using the following standard theorem (Ref. 3, p. 236): The maximum-likelihood estimators $\hat{\alpha}$ and $\hat{\nu}$ are, for large N, approximately bivariately normally distributed, with means α, ν and with covariance matrix $(1/N)$ **V**. Here, $\mathbf{V} = \mathbf{R}^{-1}$, $\mathbf{R} = -E\left((\partial^2/\partial i \partial j) \log f\ (x;\ \alpha,\ \nu)\right)$; $\partial i = \partial\alpha$ if $i = 1$, $\partial i = \partial\nu$ if $i = 2$, and E denotes expected value with respect to the distribution $F(x;\ \alpha,\ \nu)$ whose density function is $f(x;\ \alpha,\ \nu)$.

From this theorem, we could obtain joint confidence regions for $\hat{\alpha}$ and $\hat{\nu}$, but since we are interested only in $\hat{\nu}$, we proceed as follows. Since $\hat{\alpha}$ and $\hat{\nu}$ are assumed to be jointly normal, then $\hat{\nu}$ is normal in its own right. And the expected value of $\hat{\nu}$ is still ν. The variance of $\hat{\nu}$ is merely $1/N$ times the (2, 2) term in **V**. Thus, the asymptotic marginal distribution of $\hat{\nu}$ is completely specified, and we can obtain confidence intervals for ν by standard procedures. We must first, therefore, find the covariance matrix **V**.

Differentiation of Eq. 1 and 3 yields

$$\begin{aligned}
\frac{\partial^2 \log f}{\partial \alpha^2} &= \frac{-1}{\alpha^2} - (x - x_0)^2 \exp -(\alpha\,(x - x_0) + \nu)\,, \\
\frac{\partial^2 \log f}{(\partial\alpha\partial\nu)^2} &= -\,(x - x_0) \exp -(\alpha\,(x - x_0) + \nu)\,, \qquad (6) \\
\frac{\partial^2 \log f}{\partial \nu^2} &= \exp -(\alpha\,(x - x_0) + \nu)\,,
\end{aligned}$$

Defining $r_{\alpha\alpha} = -E\left(\dfrac{\partial^2 \log f}{\partial \alpha^2}\right)$, etc., we have

$$\begin{aligned}
r_{\alpha\alpha} &= \frac{1}{\alpha^2} + E\big((x - x_0)^2 \exp -(\alpha\,(x - x_0) + \nu)\big)\,, \\
r_{\alpha\nu} &= E\big((x - x_0) \exp -(\alpha\,(x - x_0) + \nu)\big)\,, \qquad (7) \\
r_{\nu\nu} &= E\big(\exp -(\alpha\,(x - x_0) + \nu)\big)\,.
\end{aligned}$$

We must now find the expected values required in Eq. 7. We define $H(t)$, the moment generating function of the standardized extreme value distribution $\exp(-\exp(-y))$, as

$$H(t) = \int_{y=-\infty}^{\infty} \exp(ty) \exp -(y) \exp -(\exp -(y))\, dy\,. \qquad (8)$$

Then $H(t) = \Gamma\,(1 - t)$, where Γ is the gamma function (Ref. 2, p. 173). We can use this result by transforming Eq. 7:

$$\begin{aligned}
&E\big((x - x_0)^2 \exp -(\alpha\,(x - x_0) + \nu)\big) \\
&\quad = \int_{y=-\infty}^{\infty} \left(\frac{y - \nu}{\alpha}\right)^2 \exp -(2y) \exp -(\exp -(y))\, dy\,, \\
&E\big((x - x_0) \exp -(\alpha\,(x - x_0) + \nu)\big) \qquad (9) \\
&\quad = \int_{y=-\infty}^{\infty} \left(\frac{y - \nu}{\alpha}\right) \exp -(2y) \exp -(\exp -(y))\, dy\,, \\
&E\big(\exp -(\alpha\,(x - x_0) + \nu)\big) \\
&\quad = \int_{y=-\infty}^{\infty} \exp -(2y) \exp -(\exp -(y))\, dy\,.
\end{aligned}$$

The third integral in Eq. 9 is easy to evaluate and equals 1. For the other two integrals, we define

$$\begin{aligned}
A_1 &= \int_{y=-\infty}^{\infty} y \exp -(2y) \exp -(\exp -(y))\, dy\,, \\
A_2 &= \int_{y=-\infty}^{\infty} y^2 \exp -(2y) \exp -(\exp -(y))\, dy\,. \qquad (10)
\end{aligned}$$

Then Eq. 9 becomes

$$\begin{aligned}
E\big((x - x_0)^2 \exp -(\alpha\,(x - x_0) + \nu)\big) &= \frac{A_2}{\alpha^2} - \frac{2\nu A_1}{\alpha^2} + \frac{\nu^2}{\alpha^2}\,, \\
E\big((x - x_0) \exp -(\alpha\,(x - x_0) + \nu)\big) &= \frac{A_1}{\alpha} - \frac{\nu}{\alpha}\,, \qquad (11) \\
E\big(\exp -(\alpha\,(x - x_0) + \nu)\big) &= 1\,.
\end{aligned}$$

Thus, we are reduced to determining A_1 and A_2. But from Eq. 8 we have

$$\begin{aligned}
A_1 &= H'\,(-1) = -\,\Gamma'\,(2)\,; \\
A_2 &= H''\,(-1) = \Gamma''\,(2)\,. \qquad (12)
\end{aligned}$$

Using Whittaker and Watson (Ref. 4, Chapter XII), we recall that

$$\begin{aligned}
\Gamma'\,(2) &= 1 - \gamma\,, \qquad (13) \\
\Gamma''\,(2) &= (\Gamma'\,(2))^2 + \frac{\pi^2}{6} - 1 = (1 - \gamma)^2 + \frac{\pi^2}{6} = 1\,,
\end{aligned}$$

where γ is Euler's constant .5772.... . Substituting these values for A_1 and A_2 into Eq. 11 and then into Eq. 7, we finally find the matrix $\mathbf{R}$:

$$r_{\alpha\alpha} = \frac{1}{\alpha^2}\left((1-\gamma)^2 + \frac{\pi^2}{6} + 2\nu(1-\gamma) + \nu^2\right),$$
$$r_{\alpha\nu} = \frac{1}{\alpha}(1-\gamma+\nu), \qquad (14)$$
$$r_{\nu\nu} = 1.$$

We are now ready to derive the covariance matrix $\mathbf{V}$, or rather its (2,2) term. Since $\mathbf{V} = \mathbf{R}^{-1}$, the (2,2) term of $\mathbf{V}$ is given by $(1/(r_{\alpha\alpha}r_{\nu\nu} - (r_{\alpha\nu}))^2)\, r_{\alpha\alpha}$. Remembering to divide by N, we have

$$\mathrm{Var}\,\hat{\nu} \approx \frac{6}{N\pi^2}\left((1-\gamma+\nu)^2 + \frac{\pi^2}{6}\right). \qquad (15)$$

Now in Eq. 15, ν is an unknown parameter. We could replace ν by its estimator $\hat{\nu}$, since N is large and therefore $\hat{\nu}$ is likely to be close to ν. However, for a more accurate estimation, we proceed as follows.

A one-sided confidence interval on the error probability is desired. That is, we wish to say: unless an event with a certain low probability has occurred, the true error probability is less than a certain value. This is the criterion upon which acceptance or rejection of a given detector under test is to be based. And a one-sided-on-the-left confidence interval for the probability that x exceeds x_0 corresponds to another confidence interval for ν, one-sided on the right.

If a confidence interval of confidence λ(λ a number slightly less than 1) is desired, we demand a ν_0 such that

$$Pr(\nu > \nu_0) = \lambda \qquad (16)$$

in the *a posteriori* sense. To do this, we seek a ν_1 such that

$$Pr(\hat{\nu} < \nu_1) = \lambda \qquad (17)$$

where probability is interpreted according to the marginal distribution of ν. Since ν has mean ν and variance σ^2 given by Eq. 15, we write Eq. 17 as

$$Pr\left(\frac{\hat{\nu}-\nu}{\sigma} < \frac{\nu_1-\nu}{\sigma}\right) = \lambda. \qquad (18)$$

Since $\hat{\nu}$ is normal, $(\hat{\nu}-\nu)/\sigma$ has the unit normal distribution. Now define Φ_λ as the quantile of order λ of the unit-variance normal distribution (that is, with probability λ, a unit normal variate is less than Φ_λ). We then have, from Eq. 16,

$$\frac{\nu_1-\nu}{\sqrt{\frac{6}{N\pi^2}\left((1-\gamma+\nu)^2 + \pi^2/6\right)}} = \Phi_\lambda, \qquad (19)$$

so that one obtains a quadratic equation for ν:

$$\nu^2\left(1 - \frac{6\Phi_\lambda^2}{N\pi^2}\right) - 2\nu\left(\nu_1 + \frac{6\Phi_\lambda^2}{N\pi^2}(1-\gamma)\right) + \nu_1^2 - \frac{6\Phi_\lambda^2}{N\pi^2}\left((1-\gamma)^2 + \frac{\pi^2}{6}\right) = 0. \qquad (20)$$

Only one of the two roots of Eq. 20 is acceptable, since ν_1 must be greater than ν when $\Phi_\lambda > 0$ (the case of interest corresponds to $\lambda > \frac{1}{2}$). Thus

$$\nu_1 = \nu + \left(\sqrt{\frac{6}{N\pi^2}\left((1-\lambda+\nu)^2 + \frac{\pi^2}{6}\right)}\right)\Phi_\lambda. \qquad (21)$$

Consequently,

$$Pr\,\hat{\nu} < \nu + \left(\sqrt{\frac{6}{N\pi^2}\left((1-\gamma+\nu)^2 + \frac{\pi^2}{6}\right)}\right)\Phi_\lambda = \lambda. \qquad (22)$$

Consider the region in which $\hat{\nu} < \nu + c\sqrt{(\nu+a)^2 + b^2}$ for a,b,c as positive constants and $c < 1$. One has

$$(\hat{\nu}-\nu)^2 < c^2\left((\nu+a)^2 + b^2\right), \qquad (23)$$

which ultimately reduces to

$$\nu > \frac{1}{1-c^2}\left(ac^2 + \hat{\nu} - c\sqrt{(\nu+a)^2 + (1-c^2)b^2}\right). \qquad (24)$$

The minus sign must be taken in Eq. 11, since the plus sign will ultimately make ν_0 greater than $\hat{\nu}$. But ν_0 must be less than $\hat{\nu}$, for $\hat{\nu}$ corresponds to a confidence of .5, whereas we are interested in larger confidences. In the problem at hand, $c = (\sqrt{6/N\pi^2})\,\Phi_\lambda$, $a = 1-\gamma$, $b = \pi^2/6$.

We now have $Pr\,(\nu > \nu_0) = \lambda$, where ν_0 is given by

$$\nu_0 = \frac{1}{1-c^2}\left(ac^2 + \nu - c\sqrt{(\nu+a)^2 + (1-c^2)b^2}\right). \qquad (25)$$

IV. COMPARISON WITH THE ERROR-COUNTING METHOD

Let us now compare the confidence intervals for e (the probability that a "1" changes to a "0") obtained using extreme-value theory, with those obtained using the binomial distribution. This latter method would ignore the structure of an error (namely, that an error is caused by the large deviation of a random quantity), and would record only the occurrence of errors. Thus, in the data of Table 1, the threshold x_0 was never exceeded, and so no errors were made. The error-counting method only uses the data "no errors in 3000 bits" instead of the actual record of voltage deviations. Let us now determine how wide the confidence interval for the error probability e is when using the binomial distribution in this way.

We have 3000 independent samples from a binomial distribution, with unknown parameter e for the probability of success. Find a 90% confidence interval for e of the form $(0, e_1)$, if no "successes" (errors) are observed. To rephrase the problem, how large can e_1 be so that the binomial distribution with parameter e_1 has probability .1 of having no successes occur in 3000 trials?

Since e_1 is small, we can use the Poisson approximation to the binomial distribution: the probability of no successes in n trials when the probability of success is e_1 is approximated, for large n and small e_1, by $\exp(-ne_1)$. Thus, $\exp(-3000e_1) = .1$, $3000e_1 = 2.3026$, $e_1 = 7.7\times10^{-4}$. Since e_0 was 6×10^{-4}, the binomial method in this instance is worse than the extreme-value method (in the sense of length of confidence interval) by a factor of 1.3, when no errors are counted.

A more dramatic advantage appears when the true e is still smaller, say 10^{-5}. Then the extreme-value method gives reasonable confidence intervals having some relation to the true e. But the binomial method keeps giving the same $e_1 = 7.7\times10^{-4}$ when no error is observed, regardless of the true e! Thus the error-count method loses all power to distinguish between error rates of 10^{-4} and 10^{-5} with 3000 samples, whereas the extreme-value theory can easily distinguish between the two error rates with only 3000 samples.

For instance, to distinguish 10^{-4} and 10^{-5} error rates at the .1 significance level with the error-count method would require a much larger number of samples. Specifically, we require that the e_1 for the .1 level, when no errors are made, be less than 10^{-4}. Let M samples be required. Then $\exp(-Ne_1) = .1$, $Me_1 = 2.3026$; but $e_1 < 10^{-4}$ so $M > 23{,}026$. Thus eight times as many samples are required by the error-count method as by the extreme-value method in this instance. The savings are even more dramatic at lower significance levels (.01 instead of .1), or at lower error rates. The reason for adopting the extreme-value theory is clear.

V. GOODNESS OF FIT

In adopting extreme-value theory, we must question two major assumptions. The first assumption is not so serious—that the value of N, the number of samples of the extreme-value distribution available for estimating the error probability e, is large enough to use the asymptotic distribution of the maximum-likelihood estimators. The N we used was only 30, so the question is indeed relevant. However, this question is attackable analytically, and is not a fundamental question on the applicability of the extreme-value distribution.

The more serious question is whether extreme-value theory holds at all. We have the large number ($n=100$) of independent samples from which to choose extremes. The n looks large enough, if the original distribution of voltage deviations in the detector is not badly behaved. But in the case of many detectors, the voltage whose distribution is in question is the output of a highly complicated nonlinear device such as a tuning-fork filter. The signal into the "front end" of the detector is a certain audio frequency if the bit is a "1," plus additive white

Gaussian noise. The output of the tuning-fork filter is detected, sampled, and compared with the threshold x_0. If x_0 is exceeded, the bit is called "1"; if x_0 is not exceeded, the bit is called "0."

Now the behavior of Gaussian processes through nonlinear operations is fairly well known qualitatively, even if the particular form of the distribution cannot be readily found; it is reasonable to expect the tails of the distribution to be of the exponential type. Consequently, extreme-value theory ought to apply.

However, we can test whether extreme-value theory holds, so as not to rely too heavily on any of the above arguments—we can apply a goodness-of-fit test to the data to which extreme-value theory supposedly applies. Such a test tells us whether it is reasonable to assume that a given empirical distribution could have arisen as a set of N independent samples from an assumed distribution.

We shall apply a certain goodness-of-fit test to the data of Table 1. The null hypothesis is that the true distribution of voltage fluctuations is a Gumbel distribution with parameters α and ν chosen as $\alpha = 8.76\times10^{-3}$, $\nu = 3.52$, the maximum-likelihood estimators found in Part II. The procedure to be described is a good one to use in testing goodness-of-fit to any Gumbel or other extreme-value distribution.

Just as in the case of confidence intervals, a one-sided goodness-of-fit test is desired. Furthermore, the test should have its high power against those alternatives that would lead us to think that the error probability is lower than it really is. Bearing these two requirements in mind, we have constructed the following goodness-of-fit test for use in determining whether Gumbel's extreme-value theory holds in a given situation. This material has previously been published (Ref. 5); a summary will be given here.

Recall that a test is called *distribution-free* if the distribution of the test statistics being used, under the null hypothesis that the assumed distribution is the true one, is independent of the assumed true distribution. The advantage of distribution-free tests is that only one table is required, instead of a different table for each null hypothesis. The test to be obtained is of this distribution-free kind.

In Ref. 6, a family of distribution-free goodness-of-fit tests is described. A test is called a one-sided *Kolmogorov–Smirnov test* if the test statistic used is of one of the two forms $\sqrt{N}\sup_x\left(\pm(F(x)-F_N(x))\right)$; where F is the assumed distribution, supposed continuous everywhere; F_N is the distribution of N samples from the actual distribution; and $\Psi(F)$ is a fixed non-negative weight function (selected according to what the alternative hypothesis to F might be). These tests are all distribution-free. In Ref. 7, Doob used the theory of Markov processes to derive the asymptotic distribution of the test statistics for large N (this explains the use of the factor $\sqrt{N}$ above), when Ψ is identically equal to 1. The answer is surprisingly simple —the asymptotic distribution of the test statistic is $1-\exp(-2t^2)$.

Since Gumbel's theory is concerned with extreme values, the alternatives to the null hypothesis that one is afraid of are deviations in the right-hand end of the distribution. One fears that the values of the maximum of the random voltage given by the true distribution are even larger than the Gumbel distribution would indicate, since then detectors would be accepted as good when in fact they yield too high an error probability. That is, the alternatives against which one wishes to have high power are of the form "the true random variable is *stochastically greater* than the sample distribution would indicate."

This suggests that the test statistic should weight values of $F(x)$ close to 1 more than it does values close to 0. However, one is also interested in whether the theory holds at all, and not only in whether the large values of x are given correct probabilities. Thus, the test statistic should also give some weight to small x. These facts suggest $\Psi(F) = F$ as the test statistic.

Define

$$P_N^+ = \sqrt{N}\sup_x\left(F(x)-F_N(x)\right)F(x)\,.$$

The "max" rather than the "min" is chosen, because when F_N is below F by a large amount, then the true values of x tend to be larger than the null hypothesis would indicate; for if fewer x are below x_0 (that is, if $F_N(x_0) < F(x_0)$), then x exceeds x_0 more than F would indicate. And the weight function $F(x)$ weights the larger x_0 more than it does the smaller. Consequently, if $F_N(x_0)$ were less than $F(x_0)$ near the left-hand tail, the statistic P_N^+ might not pick up this deviation, as it need not by the heuristic motivation for the test.

The advantage of using this special Kolmogorov–Smirnov test with weight function $\Psi(F) = F$ is that the new test

has higher power against alternatives of the form "the true distribution at a given x is less than the assumed distribution." That is, the test has higher power against alternatives of the form "the true random variable is stochastically greater than the assumed one." We gain this extra power where we need it by giving up our power against alternatives of the form "the true random variable is unequal to the assumed one."

The distribution of this new statistic for $N = 30$ was found by a Monte–Carlo method, using a scheme which gave the correct distribution in the known case of $\Psi(F)=1$, tables of which for various N appear in Ref. 8.

From Fig. 2, which graphs the data of Table 1 and Fig. 1 in more usual coordinates, we can readily find the maximum of $\sqrt{30}\,(F(x) - F_{30}(x))\,F(x)$, since the maxima can be shown to occur at the jumps of $F_{30}(x)$. The maximum deviation P^+_{30} occurs at the eleventh largest x, i.e., at $u_{(11)}$, and so $P^+_{30} = .22$.

Our calculations on the distribution of P^+_{30} show that deviations this large or larger occur with probability .72. Thus, the agreement is excellent. The ordinary two-sided Kolmogorov–Smirnov test was also applied, and gave very good agreement too. Thus, the Gumbel theory holds for this data.

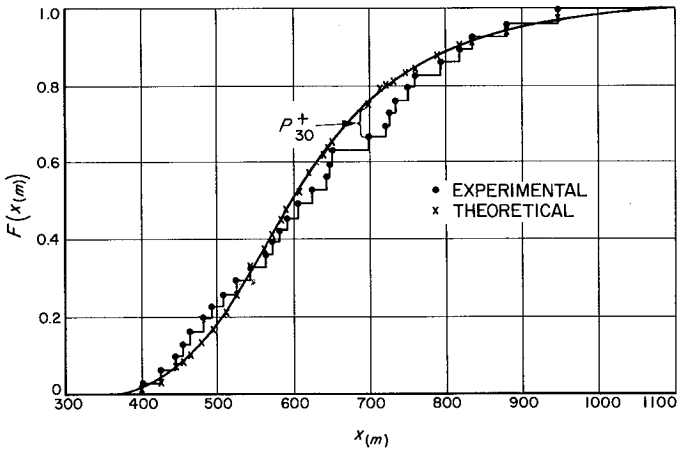

Fig. 2. The goodness-of-fit test

Another type of test was performed on the data of Table 1 to check the internal consistency of the Gumbel method. If the theory is to hold, exceedance probability estimates from a short run of data should yield estimates related to, but not as good as, the estimates obtained with further data. The following experiment was performed. The 30 extremes of Table 1 were divided into two groups of 15 each at random. One of the groups was chosen, and the maximum-likelihood Gumbel distribution was fitted on extreme-value probability paper for these 15 values. On the same paper, the line for all 30 points was drawn for comparison (see Fig. 3).

Note how close the two lines are—well within an error explainable by random deviations in sampling. The maximum-likelihood exceedance probability estimated from all 30 extremes was .024; the set of 15 gives .020. Thus, the theory is internally consistent within the data. The goodness-of-fit test, coupled with this internal consistency test, justifies the use of Gumbel's extreme-value theory in estimating error probabilities. The large savings in testing time demonstrated in Part IV are therefore real.

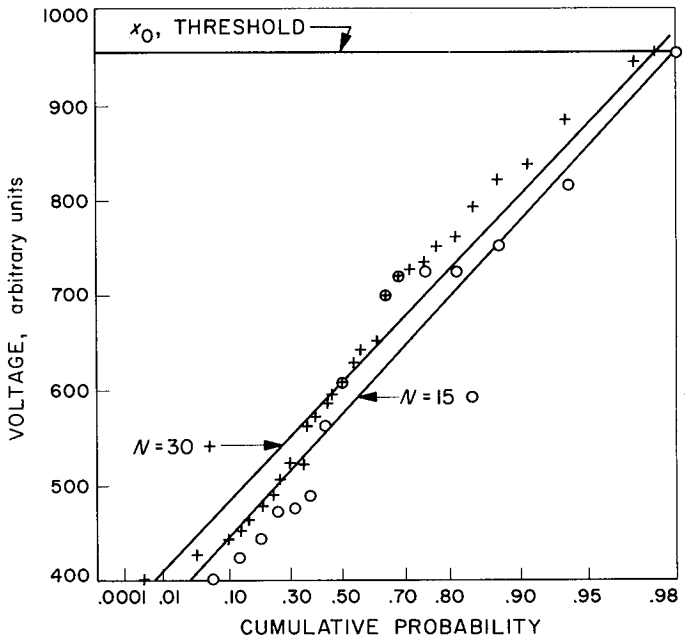

Fig. 3. Consistency in two different sample sizes